

Group Assignment Question and Guidance

---

**Review Prediction: eCommerce Platform**

**Background:**

You are working for an analytics consultancy company. A large South American eCommerce platform, “Nile”, requested you to deliver a pitch to win a major contract with them to develop and deploy a model to predict which customers will leave positive reviews.

This model will be used to then specifically target/incentivise these customers to write a review. Maintaining a set of positive reviews is very important in online retail. Being able to ensure that customers are creating these reviews is significant, but also there is a need to do so in a resource efficient way,

You are provided with an excerpt from their database system, including 8x tables of data exported as CSV. They also include a file to help translate the product categories into English. The data can be joined according to this diagram:

Your task is to appropriately join the data you wish to include (note, you do not need to include every table in your analysis) and prepare the data for modelling. Your goal is to then build a prediction model, using an appropriate machine learning algorithm, to predict customer reviews (from the “olist\_order\_reviews\_dataset”). You can choose to build a regression model (predict star ratings as numbers), a multi-class classification model (predict star ratings as multiple classes) or a binary-class classification model (predict star ratings as above or below a threshold – e.g. “1-3 stars” or “4-5 stars”).

**Presentation:**

You and your team will be required to pitch your results and recommendations in a presentation of 10 minutes to Nile’s “management board” impersonated by WBS staff. The pitch needs to be recorded and submitted as a video file. In the video, one should be able to see the slides (or other means of presenting) as well as the presenter – the latter is important since we also evaluate presentation skills. We leave it to you whether you want to have a single presenter or whether all of you would like to present; we will not penalise for changes of presenters (which in practice is not a good idea since it is apparently quite disruptive).

Keep it professional - business apparel is expected! Please note that this is not an exercise in professional video production; the evaluation will be based on the content of the presentation - to that end, make sure that you motivate the project properly such that it is very clear why we (as Nile) should invest in your services. Keep in mind in designing the presentation that the audience are your clients; they are managers who may not necessarily have technical knowledge so reflect carefully on what you present. They will be interested in the meaning of your work for their business, and whether their investment in this project will be worth it.

Your team also needs to submit a technical report in document form. Technical considerations should be described in the report, see instructions below. The report is subject to a limit of 2,000 words.

**Instructions:**

Please read the instructions carefully and discuss with your colleagues and provide an outline of your approach.

[25 marks] Data Engineering: Prepare the data for modelling. You will need to perform sensible steps to join, aggregate, clean and feature engineer the dataset.

[25 marks] Modelling: Implement a classical machine learning model (or multiple) on the provided data to identify a suitable approach for predicting reviews. The model should be justified for its use in the problem space, and you need to demonstrate sufficient and appropriate evaluation metrics.

[20 marks] Overall approach: Outline your approach considering the steps in CRISP-DM methodology and justify the selected methods for data, modelling and evaluation steps used in this project. You do not need to provide every detailed step in the report, instead, focus on the big picture and justifications of the chosen approach and the chosen evaluation measures. The focus of this assignment is to give you a go at thinking about the business problem, data preparation for modelling, implementing and interpreting several models using real data, describing your approach to the machine learning project clearly and providing clear justification of the steps followed and the methods used. You may wish to mention any considerations on deployment, but you are advised to keep these brief.

A python (.py) or notebook (.ipynb) should be submitted as well (of course without the data file).

[30 marks] Presentation: These final marks are awarded for the sales pitch, where we will evaluate presentation style, content, managerial insights, media, and appropriateness for the audience and generally how convincing your presentation is.

There is no single correct approach to this assignment; the focus is on demonstrating proper management of a realistic analytics project using the CRISP-DM methodology and that you can and suitably present managerial insights to your client.

**Submission Instructions:**

Project report and presentation files should be submitted as pdf documents.

Please check the Student Handbook – Section 6.2b Assignment Preparation for the format of the project report.

Python code (.py or .ipynb) should be submitted without the data file.

Please add your project report, presentation and python code to the same folder and bundle it as a Zip file.

The file should be named as:

*group\_number\_X.zip*

where X is your given group number.

Presentation video will be submitted separately and the instructions will be announced in due time.

**Minutes of Group Meetings:**

This is a non-assessed part of the group project. We expect all group members to contribute to the group project. Please submit your project meeting minutes (notes) and action list (tasks per individual team members) agreed by all team members. The meeting minutes should be submitted as a pdf document. Please check MeetingMinutes.pdf for instructions.

**REMEMBER THAT ALL TEAM MEMBERS MUST SUBMIT PEER ASSESSMENT MARKS BEFORE THE DEADLINE AS ADVERTISED ON MY.WBS.**

---

**Presentation Time:** Maximum 10 minutes per group (*this is a strict limit, not a guideline*); presentations should be recorded for submission

**Submission Format:** You should submit your group presentation as a recording, with narration over the slides. This can be achieved in Powerpoint using the guidance [here](#). We suggest you practice using Powerpoint if you are not familiar with any of these features. Do not provide information in the notes pages as these will not be read.

One member of each group should make the submission on behalf of the whole group.

**Assessment Weighting:** 40%

**Submission Deadline:** Monday 2 December 2024 before 12:00:00 (UK time)

**Marks Released by:** Wednesday 8 January 2024  
(we will aim to release marks by this date, but in the event of an unavoidable delay a message will be posted on the module page)

**Artificial Intelligence:** PERMITTED

**FT Courses**

**Further Information:** Guidance on “Successful Group Work” can be found in your [Student Handbook](#)

---

**Use of Artificial Intelligence**

The University recognises an increasing number of technologies such as Artificial Intelligence and that they may be applicable in your completing this assessment. The assessment brief sets out specific requirements or restrictions, and your student handbook has further guidance and advice.

You are reminded that the inappropriate use of such a technology may constitute a breach of University policy, such as the Proofreading Policy or [Regulation 11 \(Academic Integrity\)](#). If you breach these policies, it may have significant consequences for your studies. Please make sure you read and understand the assessment brief and how AI may or may not be used.

**Failure to disclose the use of any AI at the point of submission may be prejudicial in any later investigations, should they arise.**

For this assessment, AI is:

**PERMITTED (neither prohibited or required)**

If you use a generative Artificial Intelligence (AI) in the process of completing this assessment you **MUST** set out clearly the following:

- WHY you used a generative AI
- WHAT it was used for
- WHICH AI was used; and
- If any generated content has been used directly in this submission, if so where.

Note that this declaration does NOT contribute towards the word count for the assessment.

You will also have to confirm in your declaration that the work remains yours and you have intellectual ownership of it. You may be called for viva or other interview to demonstrate such intellectual ownership. A failure to disclose the use of AI, or the use of a misleading description of its use may have significant consequences for your studies. As a result, keeping good records of your interactions is strongly advised.

---

**Submitting your work**

Before you submit your assessment, you should ensure you are familiar with the guidance and rules in the “Your Assessments” section of your [Student Handbook](#), paying particular attention to:

- [Assignment Preparation](#)
- [Academic Integrity \(including Plagiarism\)](#)
- [Referencing](#)
- [Word Count Policy and Formatting](#)
- [Confidentiality of your work](#)
- [Deadlines and Extensions](#)
- [Late submission of work](#)
- [Guidelines for Online Submission](#)

---

**Mitigating Circumstances**

Mitigating circumstances **MUST** be submitted **within 20 working days following the submission deadline**. Mitigating circumstances not submitted by the relevant deadline are not required to be considered by the School/Department and may have to be considered by an Academic Appeals Committee as part of an academic appeal – for further information, please see:

<https://warwick.ac.uk/services/gov/calendar/section2/regulations/reg42academicappeals>

Please see your Student Handbook for more guidance on mitigating circumstances

---